With public health focusing on reduction of morbidity and mortality in the human population, transmission and virulence are major targets of molecular disease surveillance. Current molecular approaches to the detection of HCV transmission are based on phylogenetic analysis of intra-host HCV variants. Next-generation sequencing and mass spectrometry offer significant improvements in accuracy, throughput and cost of the transmission identification. Resistance to interferon (IFN) is an important virulence factor as it is essential for establishment of HCV infection and development of disease. IFN resistance has been shown to be associated quantitatively with epistatic connectivity among sites in the HCV genome, thus paving the way for genetic detection of virulence. Integration of genetic testing with computational models that automatically interpret the complex genetic parameters into transmission events and virulence is fundamental to comprehensive molecular surveillance of hepatitis C.

Hepatitis C is caused by HCV [1] and is recognized as one of the leading causes of chronic liver disease associated with end-stage cirrhosis and liver cancer [2]. An estimated 130 million people are HCV-infected worldwide. The prevalence of HCV infection varies in different geographic regions of the world, ranging from <1% in Northern Europe to 15–20% in Egypt [3]. In the US, the incidence rate of acute hepatitis C rapidly declined after 1989 by approximately 90% and stabilized at the level of approximately 0.7 per 100,000 individuals during 2004–2006. However, cases of acute liver disease account for approximately only 20–30% of all newly acquired HCV infections [4]. Because acute infections and less-advanced stages of chronic disease are usually asymptomatic [5], with only half of viraemic patients having elevated alanine aminotransferase (ALT) levels [6], surveillance of HCV infections in the human population can be challenging if it is exclusively based on clinical features.

Epidemiological surveillance for the prevalence and incidence of HCV infections and of hepatitis C is crucial for monitoring trends in their dissemination, and for assessment of the efficacy of control measures. The major goal of public health is reduction of morbidity and mortality in the human population. In this respect, surveillance of disease provides a direct assessment of the state of the public health. When connections between symptomatic cases and epidemiological parameters are established, disease surveillance may be used to estimate these parameters accurately. For example, population-based surveillance for acute hepatitis C was extremely successful in identifying risk factors for acquisition of infection by evaluating the frequency of exposures reported by cases compared with population-based controls [7,8], in part, because the ratio of acute cases to HCV infections was found to be constant among the types of exposure [4].

Molecular surveillance is an important adjunct to epidemiological surveillance. It is based on tracking genetic factors associated not only with HCV infection but also hepatitis C. With disease control being the major focus of public health, molecular surveillance should directly target two fundamental parameters: transmission, which is the process responsible for dissemination of HCV infection, and virulence, which is the capacity of HCV to cause disease.

Transmission

The epidemic history of hepatitis C in the 20th century is very dynamic. In the middle of that century, a rapid increase was detected in the effective number of infections, with subtypes 1A and 1B globally [9], genotype 2 and subtype 3A in Africa [10,11] and subtype 3A in Pakistan and Central Asia [11,12]. During the late 1980s to early 1990s, an equally rapid decline in the census number of acute hepatitis C was observed in the US [4].

Review

Molecular surveillance of hepatitis C

Yury Khudyakov

Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, GA, USA

*Corresponding author e-mail: yek0@cdc.gov

Introduction

Epidemiological surveillance for the prevalence and incidence of HCV infections and of hepatitis C is crucial for monitoring trends in their dissemination, and for assessment of the efficacy of control measures. The major goal of public health is reduction of morbidity and mortality in the human population. In this respect, surveillance of disease provides a direct assessment of the state of the public health. When connections between symptomatic cases and epidemiological parameters are established, disease surveillance may be used to estimate these parameters accurately. For example, population-based surveillance for acute hepatitis C was extremely successful in identifying risk factors for acquisition of infection by evaluating the frequency of exposures reported by cases compared with population-based controls [7,8], in part, because the ratio of acute cases to HCV infections was found to be constant among the types of exposure [4].

Molecular surveillance is an important adjunct to epidemiological surveillance. It is based on tracking genetic factors associated not only with HCV infection but also hepatitis C. With disease control being the major focus of public health, molecular surveillance should directly target two fundamental parameters: transmission, which is the process responsible for dissemination of HCV infection, and virulence, which is the capacity of HCV to cause disease.

Transmission

The epidemic history of hepatitis C in the 20th century is very dynamic. In the middle of that century, a rapid increase was detected in the effective number of infections, with subtypes 1A and 1B globally [9], genotype 2 and subtype 3A in Africa [10,11] and subtype 3A in Pakistan and Central Asia [11,12]. During the late 1980s to early 1990s, an equally rapid decline in the census number of acute hepatitis C was observed in the US [4].
Such dramatic variations in the size of HCV population had a major impact on HCV genetic diversity [9–11].

Patterns of transmission may differentially change in different risk groups over time. Such changes occurred during 1982–2006 in the US when injecting drug use emerged as the most commonly reported risk factor for HCV infection [4]. Although the number of injecting drug use related cases of acute hepatitis C declined after 1989, the proportion of cases actually increased from 32% during 1982–1989 to 46% during 1994–2006 [4]. With the overall prevalence of HCV infection being >80% in injecting drug users (IDUs) [13,14], this high-risk group would have been experiencing primary infection and reinfection more frequently than the general population. Such infections may not be accompanied by clinical symptoms and may be masked by symptoms arising from pre-existing chronic hepatitis C, thus preventing detection by disease surveillance. The success of frequent HCV transmissions, for example, among IDUs, generates conditions for rapid inter-host HCV evolution. In the general population, however, HCV infections are less frequent, being acquired predominantly from chronically infected individuals.

In addition to changing the pace of evolution, predominant acquisition of infection from acutely or chronically infected patients can affect the course of inter-host HCV evolution because the intra-host constitution of HCV populations found in acute and chronic stages of infection can be different. Intra-host HCV evolution is a very complex process defined by adaptation to the host and is associated with significant temporal variations in the genetic composition of HCV variants or quasispecies [15]. The genetic changes usually manifest as variation in expression of viral traits, for example, the capacity to establish infection in a new host, with HCV variants from the acute stage being potentially most infectious [16].

Variation in the rate and patterns of transmission strongly affects the genetic composition of the HCV population [9–12], which can result in phenotypic variation that affects virulence. It was observed that even minor genetic changes were associated with severity of liver disease [17,18] and the propensity of HCV to resist therapeutic treatment [19,20]. Additionally, the frequency of virus transmission can affect virulence [21]. The rate of spread of HCV increased dramatically in the mid-20th century [9–12], and is currently being maintained at a high level among, for example, IDUs [4,22]. Frequent passing of HCV from one person to another significantly relaxes the selection pressure for the establishment of chronic infection, thus potentially decreasing the HCV dependence on low virulence during the late stages of infection in order to sustain viral circulation among hosts. It is therefore important to investigate the association between the rate of transmission and virulence.

Detection of transmission is essential for tracking HCV infections and hepatitis C. Transmission is a fundamental viral process responsible for dissemination of infection and disease. Surveillance based on the manifestation of the disease is limited in its capacity to track transmissions. Viral genetic testing is a more efficient way of proving direct transmissions. It involves assessment of genetic relatedness among HCV strains and is based on the assumption that the HCV strain shared by two patients has identical or very similar genetic composition in both patients. Accurate evaluation of genetic relatedness among HCV strains can be achieved using a sample of sequences of short genomic regions obtained from the intra-host viral population [23]. Although analysis of quasispecies is very complex and labour-intensive, it cannot be replaced by simple consensus sequencing because consensus sequences of short regions do not represent the intra-host HCV population adequately [23]. Furthermore, intra-host HCV evolution results in considerable changes in the composition of HCV population in patients over time [15]. The HCV genome contains hypervariable region 1 (HVR1) located at the 5’-end of the E2 gene. Owing to its variability, HVR1 is frequently used for identification of HCV transmission [23–25]. Analysis of quasispecies involves separation of individual intra-host HVR1 variants by genetic cloning [26] or by end-point limiting dilution (EPLD) real-time PCR [23] followed by sequencing. The size of the sample is usually limited to 10–50 sequences, but can be extended to a few hundred [23]. Genetic cloning- and EPLD-based molecular approaches are costly and labourious. Both are based on elaborate genetic analyses of a large number of sequences and require substantial specialized expertise for interpreting molecular data to inform on transmission events. The complexity, expense and intensity of these approaches make them applicable more to investigating individual outbreaks rather than population-based surveillance.

Next-generation sequencing (NGS) combines separation of individual DNA molecules and sequencing in a single process, which significantly simplifies large-scale sampling of intra-host variants [27,28]. However, the high rate of sequencing errors generated per DNA read [29] and the biased representation of intra-host viral heterogeneity [30] hinders its adoption for the detection of transmission. Application of NGS and mass spectrometry (MS) to genetic analysis of viral populations offers a more efficient approach to the detection of transmission [31]. Although EPLD-based sequencing, NGS and MS-based analysis of HVR1 are equally accurate in the identification of genetic relatedness among HCV strains [30], NGS and MS are more applicable for high-throughput
tracking of viral transmission. With further development in computational interpretation of the complex genetic information into transmission, both NGS and MS can be completely suitable for routine and comprehensive molecular surveillance of HCV infections.

Virulence

The complex association between HCV infection and hepatitis C confounds understanding of molecular epidemiological mechanisms affecting distribution of hepatitis C in the population. Surveillance using only clinically presented HCV infections does not take into consideration host and HCV heterogeneity. HCV is genetically variable and classified into six genotypes and numerous subtypes [32]. HCV genotypes were found to be associated with severity of disease. For example, genotype 1 was shown to be associated with a more aggressive liver disease and a high risk of cirrhosis and hepatocellular carcinoma (HCC), whereas patients infected with genotype 3 tended to have steatosis and fibrosis [33]. Additional data suggest potential genotype-specific variation in the spontaneous clearance rate [34]. With HCV genotypes and subtypes having a distinct geographic distribution [35], the association between presentation of disease and infection with different HCV genotypes should be taken into consideration.

Minor genetic variations in the HCV genome were also shown to be associated with severity of liver disease. For example, the Y164F substitution in the core protein of genotype 3a was found to correlate with steatosis [36,37], while substitutions at amino acid positions 70 and 91 of the genotype 1b core were associated with development of severe insulin resistance [38]. Several nucleotide mutations in the HCV genotype 1b core gene were significantly related to increased risk of HCC [39]. Furthermore, the number of substitutions in the interferon (IFN) sensitivity-determining region of NS5A inversely correlated with mean viral load and directly with mean serum ALT level [40].

Host genetics is another important factor affecting susceptibility to HCV infection, viral clearance and severity of disease. Patients bearing haplotype C of the HAVCR1 gene were found to be differentially susceptible to infection with different HCV genotypes [41]. The IL28B CC genotype was shown to be associated with more frequent infections with HCV genotype 3 than with genotypes 1 or 4 in HIV-coinfected patients with chronic hepatitis C, indicating variation in protective effects of IL28B against infections with different HCV genotypes [42]. The presence of the IL28B CC genotype and the HLA DQB1*0301 haplotype correlated with natural clearance of HCV acute infection [43]. IL28B variability was also shown to be linked to severity of hepatitis C [44–46]. Moreover, this association was HCV genotype-specific; despite more pronounced liver pathology found in patients chronically infected with HCV genotype 3, no such association was observed for patients infected with HCV genotype 2 [46].

Viral and host genetic factors play important roles in determining virulence. There is no molecular surveillance for HCV virulence because there are no assays for its detection. Considering that virulence is a phenotypic trait encoded in viral genome, it is essential to establish the quantitative association between HCV genetic heterogeneity and the capacity to cause disease. Although such association has not been established, recent developments in detection of resistance to IFN-based anti-HCV therapy provide an indication as to how virulence may be identified.

HCV genetic composition is associated strongly with IFN resistance. Patients infected with HCV genotype 2 have been reported to achieve sustained virological response to the combined IFN and ribavirin therapy 20–40% more frequently than patients infected with HCV genotype 1 [47–49]. Genetic variations within the core, E2, NS2, P7, NS5A and NS5B proteins were reported to be related to IFN resistance [50–56]. Host genetic polymorphisms, for example, in the IL28B locus, were also implicated in IFN resistance [56–58]. It is important to note a significant concordance in genetic factors affecting HCV clearance, severity of liver disease and response to IFN-based therapy [17,38–40,44,57–61], which suggests a strong link between virulence and IFN resistance. HCV variants that resist IFN treatment seem to interfere efficiently with IFN pathways and reduce its anti-proliferative activity, thus potentially contributing to carcinogenesis [62,63]. Successful establishment and maintenance of infection is defined largely by viral interactions with the host innate immunity and, in particular, with IFN. Since chronic HCV infection is associated with development of liver disease and HCC [2], resistance to IFN is effectively an essential determinant of virulence.

Despite the established association between HCV genetic heterogeneity and outcomes of IFN-based therapy [50–52,54–56,60], genetic analyses have failed to identify a single mutation responsible for IFN resistance. Recently, however, examination of co-evolution among HCV sites showed that genome-wide coordination among substitutions is strongly linked to IFN response [64,65]. The modelled epistatic connectivity among HCV sites can be used as a complex genetic marker of resistance [66]. Owing to extensive co-evolution across the entire HCV genome, many genomic regions were found to reflect selection pressures related to IFN action, with some short regions, for example, HVR1 and a sub-region of NS5A, having a more pronounced association with susceptibility to IFN-based
therapy [64]. Sequences of the entire HCV genome [64] and individual regions [64,66] predicted outcomes of IFN/RBV therapy with 83–90% accuracy. These findings indicate that consideration of coordination among sites allows for the effective application of the specially selected group of polymorphic sites derived from short genomic regions to detect IFN-resistant HCV strains. Interpretation of epistatic connectivity into IFN resistance requires the use of complex mathematical models. Once such models are developed, genetic assays for the detection of IFN resistance should not be any more complicated than amplifying and sequencing a short HCV genomic region, the primary structure of which can be automatically interpreted to detect the extent of resistance [64,66,67].

Conclusions

Surveillance of hepatitis C is based on detection of transmission and virulence. Current molecular approaches to the detection of HCV transmission are costly and labour-intensive and require expert knowledge for interpreting molecular data. Application of NGS and MS to the analysis of intra-host HCV populations offers a significant improvement in throughput and reduction in cost. There are currently no assays for measuring HCV virulence. Recent developments in identification of IFN resistance, however, suggest that assessment of virulence can be achieved using quantitative associations between epistatic connectivity among HCV genomic sites and liver disease. Integration of complex mathematical models, which interpret complex genetic markers into epidemiologically relevant information, with genetic testing opens novel opportunities for development of simple-to-operate assays for accurate, rapid, cost-effective and high-throughput detection of transmission and virulence. Comprehensive molecular surveillance of hepatitis C, when achieved, should aid significantly in efforts to reduce morbidity and mortality associated with HCV infection.

Disclosure statement

The author declares no competing interests.

References


